

Conference Abstract

The Microbial Antarctic Resource System: Integrating discoverability and preservation of environmentally-annotated microbial 'omics data

Maxime Sweetlove, Yi Ming Gan[‡], Alison Murray[§], Anton Van de Putte[‡]

[‡] Royal Belgian Institute of Natural Sciences, Brussels, Belgium

[§] DRI, RenoReno, United States of America

Corresponding author: Maxime Sweetlove (msweetlove@naturalsciences.be)

Received: 19 Jun 2019 | Published: 10 Jul 2019

Citation: Sweetlove M, Gan YM, Murray A, Van de Putte A (2019) The Microbial Antarctic Resource System: Integrating discoverability and preservation of environmentally-annotated microbial 'omics data. Biodiversity Information Science and Standards 3: e37499. <https://doi.org/10.3897/biss.3.37499>

Abstract

Microbial organisms - including Archaea, Bacteria and unicellular Eukaryota - collectively dominate the Earth in terms of bio- and functional diversity. Their study, often constrained by technology, has strongly benefited from the recent advancements in high-throughput DNA sequencing techniques. The vast amounts of microbial data generated in the wake of these developments, however, remains severely underrepresented on open access biodiversity data repositories (e.g. the Global Biodiversity Information Facility; GBIF). Moreover, when sequencing data has been made publicly available, is often poorly annotated with metadata and environmental variables, making it difficult to find or query. Therefore, the microbial Antarctic Resource System (mARS) aims to fill this lacuna by documenting and geo-referencing microbial datasets and linking the sequence data in the International Nucleotide Sequence Database Collaboration (INSDC) repositories with the associated environmental measurements on mARS, which is aimed to be interoperable with both INSDC and GBIF. This way, mARS helps to preserve environmental data and the metadata that is crucial for the correct processing and interpretation of sequence data, while it also connects researchers via its webportal to the existing wealth of molecular information, and allows these datasets to be more effectively accessed. Given the general complexity of microbial ecological datasets, mARS needs to operate between different data

archiving standards, such as MIxS (see <https://press3.mcs.anl.gov/genosc/mixs/>), which is oriented towards DNA sequence data, and the biodiversity-based DarwinCore standard.

Currently, mARS tries to address the challenges of integrating microbial data with these existing systems as well as connecting with the communities behind them, by documenting the datasets on GBIF's extensions or investigate the feasibility of routinely processing raw sequence data into occurrence datasets using the open computing facilities offered by the European Molecular Biology Laboratory's (EMBL) MGnify resource.

Keywords

data archiving, Antarctica, microorganisms, Bacteria, Archaea, Eukaryota, MIxS, DarwinCore

Presenting author

Maxime Sweetlove

Presented at

Biodiversity_Next 2019

Hosting institution

Royal Belgian Institute of Natural Sciences